Kevin Romero Peces-Barba
kromero8@ucsc.edu

CMPS 142 - Spring 2018

Problem Set 0

1. Assume X is distributed as a Normal distribution with *mean* $= \mu$ and *variance* $= \sigma^2$. Write the probability density function for X.

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

2. Assume that the probability of obtaining heads when tossing a coin is λ.

(a) What is the probability of obtaining the first head at the (k + 1)-th toss?

The probability of obtaining a tail is $(1-\lambda)$. To get a head in the (k + 1)-th toss, all k first tosses must be tails. The probability of this happening is $(1-\lambda)^k$.
Finally, the probability of obtaining a head after k tails is $(1-\lambda)^k\lambda$.

(b) What is the expected number of tosses needed to get the first head?

The probability to get a head is λ, so at least $\lceil \frac{1}{\lambda} \rceil$ tosses are needed to get the first one.

3. Assume X is a random variable.

(a) We define the variance of X as: $Var(X) = E[(X - E[X])^2]$. Prove that $Var(X) = E[X^2] - E[X]^2$.

$$Var(X) = E[(X - E[X])^2] = E[X^2 - 2XE[X] + E[X]^2]$$

Since $X = E[X]$, then:

$$Var(X) = E[X^2 - 2E[X]^2 + E[X]^2] = E[X^2 - E[X]^2]$$

Having $E[a-b] = E[a] - E[b]$, then:

$$Var(X) = E[X^2] - E[E[X]^2]$$

Again, $E[X] = X$, so:

$$Var(X) = E[X^2] - E[X]^2$$

(b) If $E[X] = 0$ and $E[X^2] = 1$, what is the variance of X? if $Y = a + bX$, what is the variance of Y?

$$Var(X) = E[X^2] - E[X]^2 = 1 - 0^2 = 1$$

$$Y = a + bX$$
$$\rightarrow E[Y] = a + bE[X]$$
$$\rightarrow Var(Y) = E[((a+bX) - (a + bE[X]))^2] = E[(bX - bE[X])^2] = E[b^2(X - E[X])^2]$$
$$\rightarrow Var(Y) = b^2E[(X - E[X])^2] = b^2Var(X)¶$$

4. Let $f(x, y) = 3x^2 + y^2 - xy - 11x$

(a) Find $\frac{\delta f}{\delta x}$, the partial derivative of f with respect to x. Also find $\frac{\delta f}{\delta y}$.

$\frac{\delta f}{\delta x} = 6x - y - 11$ $\qquad$ $\frac{\delta f}{\delta y} = 2y - x$

(b) Find $(x, y) \in R^2$ that minimizes f.

First, take $D = \frac{\delta^2 f}{\delta x^2}\frac{\delta^2 f}{\delta y^2} - \frac{\delta^2 f}{\delta x \delta y}$, where

$$\frac{\delta^2 f}{\delta x^2} = 6$$
$$\frac{\delta^2 f}{\delta y^2} = 2$$
$$\frac{\delta^2 f}{\delta x \delta y} = -1$$

Then,

$$\frac{\delta f}{\delta x} = 0 \to y = 6x - 11$$
$$\frac{\delta f}{\delta y} = 0 \to 2y = x$$

Solving these two-equation system:

$$2(6x - 11) = x \to 12x - 22 = x \to 11x = 22$$
$$x = 2, \ y = 1$$

For this to be a minima, $D > 0$ must be true. Since $D = 6 * 2 - (-1) = 12 + 1 = 13$, $(x, y) = (2,1)$ minimizes f.

5. One way to define a *convex* function is as follows. A function *f(x)* is convex if

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

for all x, y and $0 < \lambda < 1$.

(a) Prove that $f(x) = x^2$ is a convex function.

$f(\lambda x + (1 - \lambda)y) = (\lambda x + (1 - \lambda)y)^2$, $f(x) = x^2$ and $f(y) = y^2$

$$\begin{aligned}
(\lambda x + (1 - \lambda)y)^2 &\leq \lambda x^2 + (1 - \lambda)y^2 \\
\lambda^2 x^2 + 2\lambda(1 - \lambda)xy + (1 - \lambda)^2 y^2 &\leq \lambda x^2 + y^2 - \lambda y^2 \\
\lambda^2 x^2 + 2\lambda(1 - \lambda)xy + y^2 - 2\lambda y^2 + \lambda^2 y^2 &\leq \lambda x^2 + y^2 - \lambda y^2 \\
\lambda^2 x^2 + 2\lambda(1 - \lambda)xy - \lambda y^2 + \lambda^2 y^2 - \lambda x^2 &\leq 0 \\
(\lambda^2 - \lambda)x^2 + (\lambda^2 - \lambda)y^2 + 2(\lambda - \lambda^2)xy &\leq 0 \\
(\lambda^2 - \lambda)x^2 + (\lambda^2 - \lambda)y^2 - 2(\lambda^2 - \lambda)xy &\leq 0 \\
(\lambda^2 - \lambda)(x^2 + y^2 - 2xy) &\leq 0 \\
(\lambda^2 - \lambda)(x - y)^2 &\leq 0
\end{aligned}$$

Having $(x - y)^2 \geq 0 \ \forall \ x, y$ and $\lambda \in [0,1] \to \lambda^2 <= \lambda \to (\lambda^2 - \lambda) \leq 0$. So the inequality is true.¶

(b) A n-by matrix A is a *positive semi-definite* matrix if $x^T A x \geq 0$, for any $x \in R^n$ $s.t.$ $x \neq 0$. Prove that the function $f(x) = x^T A x$ is convex if A is a positive semi-definite matrix. Note that x is a vector here.

$$f(\lambda x + (1-\lambda)y) = (\lambda x + (1-\lambda)y)^T A(\lambda x + (1-\lambda)y)$$
$$f(x) = x^T A x$$
$$f(y) = y^T A y$$
$$\beta = 1 - \lambda$$

$$(\lambda x + \beta y)^T A(\lambda x + \beta y) \leq \lambda x^T A x + \beta y^T A y$$
$$(\lambda x^T + \beta y^T)A(\lambda x + \beta y) \leq \lambda x^T A x + \beta y^T A y$$
$$(\lambda x^T A + \beta y^T A)(\lambda x + \beta y) \leq \lambda x^T A x + \beta y^T A y$$

$$\lambda x^T A \lambda x + \lambda x^T A \beta y + \beta y^T A \lambda x + \beta y^T A \beta y + \leq \lambda x^T A x + \beta y^T A y$$
$$\lambda^2 x^T A x + \lambda \beta x^T A y + \lambda \beta y^T A x + \beta^2 y^T A y + \leq \lambda x^T A x + \beta y^T A y$$
$$\lambda^2 x^T A x - \lambda x^T A x + \lambda \beta x^T A y + \lambda \beta y^T A x + \beta^2 y^T A y - \beta y^T A y \leq 0$$
$$(\lambda^2 - \lambda)x^T A x + (\beta^2 - \beta)y^T A y + \lambda\beta(x^T A y + y^T A x) \leq 0$$

$x^T A x, y^T A y, x^T A y, y^T A x$ are all scalars and ≥ 0
while $(\lambda^2 - \lambda), (\beta^2 - \beta)$ are ≤ 0, making the above true.

6. Why are you taking this class? What do you expect to learn in this course?
I'm mostly taking this class because of my interest in neural networks, and that's the point I mostly expect to learn about in the course.